

基于最小二乘法威布尔溶出曲线参数估计中的自相关性与异方差性问题探讨

邱家学*, 吴晓明#(中国药科大学, 南京 210009)

中图分类号 R9-39 文献标志码 A 文章编号 1001-0408(2013)01-0037-03
DOI 10.6039/j.issn.1001-0408.2013.01.12

摘要 目的:探讨药物累计溶出百分比威布尔(Weibull)参数估计中使用普通最小二乘法(OLS)的合理性。方法:选取溶出度数据,利用杜宾-瓦尔森(DW)检验法、斯皮尔曼(Spearman)等级相关系数检验法等对自相关性和异方差性进行分析。结果:根据数据的DW检验法结果,相关系数结果均表明在采用OLS估算Weibull参数时,确实可能存在自相关性与异方差性现象。结论:在使用OLS进行Weibull参数估计时一定要进行自相关性和异方差性检验,只有在不存在自相关性和异方差性时才有可能采用OLS进行Weibull参数估计。

关键词 威布尔参数;估计;最小二乘法;自相关性;异方差性

Exploration of Autocorrelation and Heteroscedasticity in Weibull Dissolution Parameter Estimation Base on OLS

QIU Jia-xue*, WU Xiao-ming#(China Pharmaceutical University, Nanjing 210009, China)

ABSTRACT OBJECTIVE: To explore the rationality of Ordinary Least Square (OLS) used in Weibull parameter estimation of drug accumulative dissolution percentage. METHODS: DW test and Spearman's rank correlation test were applied for autocorrelation and heteroscedasticity of dissolution data. RESULTS: Results of DW test and related coefficient of dissolution data indicated that there were autocorrelation and heteroscedasticity phenomena when OLS was used for Weibull parameter estimation. CONCLUSIONS: The test of autocorrelation and heteroscedasticity must be paid attention when OLS is used. Only when no autocorrelation or heteroscedasticity exists, can OLS be used for Weibull parameter estimation.

KEY WORDS Weibull parameter; Estimation; Ordinary Least Square; Autocorrelation; Heteroscedasticity

已有文献^[1-4]介绍了药物累计溶出百分比威布尔(Weibull)的参数估计方法,其中使用最多的是普通最小二乘法(OLS)。OLS进行参数估计时思路明确、易于理解、操作简便、易于实现,但在使用时必须符合高斯-马尔柯夫条件^[5],即如果存在自相关性、异方差性,就破坏了OLS的使用条件,据此所估计的有关Weibull参数是不正确的。为此,需要探讨药物累计溶出百分比Weibull参数估计中使用OLS的合理性。

残差是指应变量的实际测量值与估计值之差。自相关性是指残差的序列相关,异方差性是指残差的方差不尽一致。自相关性、异方差性的检验方法^[6-8]有很多,本文对药物累计溶出百分比Weibull参数估计采用经典或计算简单的杜宾-瓦尔森(DW)检验法、斯皮尔曼(Spearman)等级相关系数检验法进行分析。自相关性、异方差性检验的基本方法是:首先基于Weibull分布函数式应用OLS进行参数估计;用求出的参数代入函数式中,求出估计值;根据实际值与估计值得到残差序列;基于残差序列,再利用有关检验法等进行检验。以下将进行具体介绍。

1 药物累计溶出百分比Weibull参数估计的OLS方法

以具体数据为例进行分析,表1是文献^[9]的“KCZ微囊在SGF(pH=1.2)中的溶出度试验结果(%)”的原始数据。

* 教授,硕士研究生导师,博士。研究方向:社会管理药学。电话:025-86185009。E-mail: qjx@cpu.edu.cn

通信作者:教授,博士研究生导师。研究方向:社会管理药学。电话:025-83271555。E-mail: xmwu@cpu.edu.cn

表1 KCZ微囊的溶出度数据(%)

Tab 1 Dissolution rate of KCZ micro-capsules (%)

批号	t, min												
	5	10	15	20	30	45	65	90	120	150	180	210	270
020302a	20.90	32.40	39.80	41.30	53.10	64.90	73.70	84.03	89.94	87.00	89.94	89.90	88.40
020302b	26.70	43.00	48.80	56.30	73.50	77.20	84.05	88.80	89.12	89.83	90.36	91.22	91.60
020301a	15.42	26.24	36.60	44.70	54.04	67.92	75.44	81.77	81.77	81.79	81.86	82.16	82.89

采用OLS对表1的数据进行计算,具体思路是:

(1) 现有时间变量 t 和累计溶出百分度 F 。Weibull 分布的

$$\text{函数式为 } F(t) = 1 - e^{-\frac{(t-a)^m}{\beta}}$$

(2) 对 Weibull 分布函数式进行变换, 比如有:

$$\ln[-\ln(1-F)] = m \ln(t-a) - \ln\beta \dots\dots\dots (1)$$

(3) 赋予 a 由最小可能的值(如 0.01)一直到最大可能的值(如 $t-0.01$, 本例为 4.99)进行 OLS 计算, 从中找出使得变换式相关系数 r 绝对值最大时的各个参数(如 r, a, m, β)。

采用 OLS 进行 Weibull 参数估计的计算结果见表 2。

表2 采用 OLS 计算 Weibull 参数结果

Tab 2 Calculation results of Weibull parameter based on OLS

批号	参数			
	r	a	m	β
020302a	0.985 710	2.076 792	0.556 863	7.978 338
020302b	0.984 109	3.867 668	0.416 486	3.477 044
020301a	0.978 902	4.041 282	0.468 558	6.067 223

将计算出的各参数带入到变换式(1)中的右边,从而计算出各时间点上的 $m \ln(t-a) - \ln\beta$ 之值,以此作为 $\ln[-\ln(1-F)]$ 的估计值;再用测量数据计算出相应的 $\ln[-\ln(1-F)]$,作

为实际值。计算残差(e_i),即计算实际值与估计值之差。残差序列见表3。

表3 基于OLS的残差序列

Tab 3 Residual series based on OLS

批号	t, min												
	5	10	15	20	30	45	65	90	120	150	180	210	270
020302a	0.028 9	-0.013 5	-0.026 6	-0.160 2	-0.055 5	0.029 2	0.059 6	0.190 7	0.251 9	0.007 3	0.022 9	-0.065 6	-0.269 1
020301b	0.025 2	-0.085 2	-0.158 8	-0.101 0	0.170 8	0.089 2	0.140 6	0.173 9	0.062 6	-0.003 1	-0.057 7	-0.084 1	-0.172 5
020301a	0.035 7	-0.223 0	-0.104 8	-0.018 6	0.025 2	0.191 7	0.216 4	0.247 9	0.107 6	0.000 4	-0.084 9	-0.148 9	-0.244 8

在残差序列基础上,可以进行自相关性和异方差性的检验。

2 自相关性的DW检验

自相关分为一阶自相关和高阶自相关,一阶自相关又可分为正自相关和负自相关等。这里只进行一阶自相关分析。DW检验可用于一阶自相关检验,但不能进行高阶自相关检验。

在残差基础上可求出DW统计量,其计算公式为:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \dots\dots\dots (2)$$

查DW检验表,在 $\alpha=0.05$ 显著水平上,有2个临界值dl、du。如果计算出的DW统计量小于dl,则存在一阶正自相关;如果计算出的DW统计量大于4-du,则存在一阶负自相关。

对本例,查表得dl=1.01,du=1.34。3个批号的DW统计量分别为0.797 6、0.764 6和0.554 6,均小于dl,说明这3组数据进行OLS参数估计时均存在一阶正自相关性。

如果存在自相关性,就不能采用OLS。也就是说,在药物累计溶出百分比Weibull参数估计中,使用OLS时一定要保证不存在自相关性。由于本例已证明存在自相关性,则表1中数据是不能用OLS进行Weibull参数估计的。

3 异方差性的检验

3.1 斯皮尔曼等级相关系数检验

斯皮尔曼检验法适用于2个等级序列的相关分析。斯皮尔曼等级相关系数 r_s 的计算公式为:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \dots\dots\dots (3)$$

用于异方差性检验的具体步骤为:①作OLS计算,求出 e_i ;②取 e_i 的绝对值 $|e_i|$,将OLS计算时的自变量和 $|e_i|$ 之值分别按递增的次序求出各自的等级序列,计算对应的等级差 d_i ,再按照计算公式求出等级相关系数 r_s ;③利用等级相关系数进行显著性检验:在 $\alpha=0.05$ 显著水平上,查斯皮尔曼等级相关系数检验表,得临界值 $r_s(\alpha)$ 。如果 $|r_s| > r_s(\alpha)$,则存在异方差性。

由于在已发表的文献中未查询到可直接进行异方差性分析的原始数据,则采用模拟数据进行运算以说明可能存在异方差性现象。以表1中批号为“020301a”的原始数据为基础,以不超过10%的随机波动形成模拟数据,其中一组溶出度试验模拟数据如表4前2列所示。

对表4前2列的模拟数据,采用Excel的“规划求解”功能、以OLS原理进行计算,得Weibull参数估计值为 $\alpha=3.319 82$, $m=0.544 21$, $\beta=8.076 53$ 。

采用斯皮尔曼等级相关系数法进行异方差性检验。在OLS计算结果基础上,计算出 e_i 及其 $|e_i|$ (表4的第3、4列)、时间及 $|e_i|$ 序列的等级(表4的第5、6列)、等级差序列(表4的第7

表4 溶出度模拟数据及斯皮尔曼检验法的异方差检验过程
Tab 4 The simulation data of dissolution test and the process of heteroscedasticity test based on Spearman's rank correlation test

t, min	溶出度模拟数据, %	e_i	$ e_i $	时间值排序(升序)	$ e_i $ 排序(升序)	等级差
5	15.73	0.041 3	0.041 3	1	6	-5
10	25.62	-0.162 0	0.162 0	2	12	-10
15	32.96	-0.165 2	0.165 2	3	13	-10
20	45.99	0.072 9	0.072 9	4	7	-3
30	57.41	0.143 5	0.143 5	5	11	-6
45	65.27	0.115 0	0.115 0	6	9	-3
65	73.65	0.133 7	0.133 7	7	10	-3
90	75.81	0.010 7	0.010 7	8	2	6
120	79.62	-0.037 0	0.037 0	9	5	4
150	83.74	-0.028 8	0.028 8	10	4	6
180	87.64	0.010 5	0.010 5	11	1	10
210	89.00	-0.020 6	0.020 6	12	3	9
270	90.07	-0.113 9	0.113 9	13	8	5

列),从而计算出等级相关系数 $r_s = -0.598 9$, $|r_s| > r_s(0.05, 13) = 0.554 9$ 。所以,表4的模拟数据利用OLS进行分析时是存在异方差性的,据此计算出的Weibull参数是不正确的。

3.2 帕克检验法

帕克检验法的步骤为:①建立被解释变量和解释变量的回归方程,然后计算 e_i ;②假定异方差形式为: $e_i^2 = \alpha x_i^\beta$,令 $z_i = \ln(e_i^2)$, $w_i = \ln(x_i)$,得到线性模型: $z_i = \ln(\alpha) + \beta w_i$;③对模型的参数进行t检验,如果 β 不显著,则原模型不存在异方差;如果 β 显著,则说明存在这种形式的异方差。

对于本例, e_i 及解释变量的数据列在表4的第1、3列中,进行帕克检验法的第2、3步。经回归分析发现, β 不显著。此时,似乎应该可以得到“不存在异方差”的结论。但是,这里的“不存在异方差”是针对“原模型”而言的,原模型就是 $e_i^2 = \alpha x_i^\beta$ 。也就是说,帕克检验法是通过假定了一种异方差形式进行检验,如果没有检验出异方差的存在,并不能说明一定不存在异方差,只能说明这种形式的模型不存在异方差,其他形式不能判定;但是,只要这种形式被分析出存在异方差,那就不必再用其他形式进行判定了。对于前面模拟数据的例子,基于传统帕克检验法的形式是不能判定异方差的存在性。

4 分析与讨论

对药物累计溶出百分比Weibull参数估计,一旦出现了自相关性或异方差性,OLS就失去了其理论基础,就不可再被应用了。

实际上现有的许多软件在使用OLS时,在计算结果中会给出或多或少的相应的自相关、异方差检验统计量,但这些软件针对Weibull参数估计进行自相关或异方差检验时可能存在一定的不足,应该引起研究者的注意。(1)Excel软件提供了“回归”计算功能,其本身没有相应自相关性、异方差性检验统计量的计算。(2)SPSS软件在进行“线性回归”计算时,可以提供DW检验值,但不能进行异方差性检验。进行“线性回归”计算时还面临着一个问题,即其只能对1组数据进行计算,无法从中选优,而类似表1的数据仅1个批号就有成百上千组数据要进行计算,因为要对每一组数据进行回归计算,比较相关系数的绝对值并且从中选优,这时,SPSS就无能为力了。SPSS的“非线性回归”可以用来直接拟合Weibull分布中的3个参数^[10],虽然解决了“选优”的问题,但是不能给出有关自相关性、异方

差异性检验统计量。(3)SAS软件在进行“回归分析”时虽然可以提供自相关、异方差性检验统计量,但无法解决类似SPSS“从中选优”的问题;如果采用“非线性回归”方法,又不能提供自相关性、异方差性检验统计量。(4)EViews软件是Econometrics Views的缩写,通常称为计量经济学软件包。EViews虽然可以提供更多的自相关、异方差检验方法^[1],但面临的问题与SAS、SPSS类似。(5)MATLAB软件是为科学研究、工程设计以及必须进行有效数值计算的众多科学领域提供全面解决方案的高科技计算软件,其在数值分析、非线性动态系统的建模和仿真等方面具有强大功能,但在统计计算方面明显弱于SPSS、SAS等,可以完成有关参数求解^[2],却无法提供自相关性、异方差性检验统计量。(6)MINITAB软件是现代质量管理统计的领先者、全球六西格玛实施的共同语言,以无可比拟的强大功能和简易的可视化操作深受广大质量学者和统计专家的青睐,但其面临的问题与SPSS、SAS、EViews一样^[3]。综上所述,现有软件在自相关性、异方差性检验时均不能轻易完成,只可以采用组合方式实现。如先借助软件提供的非线性回归或曲线拟合功能,求出最佳的参数组合,再用这个最佳参数代入函数式中,利用线性回归功能进行自相关性、异方差性检验。当然,在计算出 e_i 后,还可以通过编写计算机程序实现自相关性、异方差性的检验。

从自相关性检验方法来看,几乎所有的资料都要介绍到DW检验,有的文献^[4]甚至只介绍DW方法。由此可见,DW检验方法是自相关分析中使用十分普遍且经典的方法。

从异方差性检验方法来看,从一些教科书和文献中可以见到十几种检验方法,常用的方法有斯皮尔曼等级相关系数法、帕克法、戈里瑟法、戈德菲尔特-夸特法、怀特法、布鲁尔什-培甘法、巴特利特法等^[5]。很难说哪一种方法是最好的,因为每种方法都有各自的使用条件和作用特点、优势。前述方法中的后5种均有大样本的要求^[6],虽然文献中没有指出大样本的具体标准,但是按照统计学的基本理论,样本数小于30是一定不能称作为大样本的。在前面异方差数据分析的例子中,由于其样本点为13,是不能作为大样本处理的。相对而言,等级相关系数法、帕克法没有类似的限制要求。帕克检验法受到异方差形式的影响,按照帕克检验得出不存在异方差性的结论,只是对特定函数形式而言,在采用其他函数形式的假定下仍有可能存在异方差。斯皮尔曼等级相关系数法原理简单、计算方便,当然还有着一定的不足,如易受到模型异常值点等特殊样本数据的影响^[7]。斯皮尔曼等级相关系数法实际上是一种非参数检验法。非参数检验法与参数检验法相比,犯第一类错误的概率小,犯第二类错误的概率大。也就是说,尽管斯皮尔曼等级相关系数法未检验出异方差的存在,但不能肯定异方差就一定不存在;反之,只要检验出存在异方差,其结论就不容置疑了。

此外,自相关性、异方差性的检验方法与高斯-马尔柯夫条件并不矛盾,只是用于保证满足高斯-马尔柯夫条件的检验。

当存在自相关性、异方差性时,就应该使用其他不包含OLS法思想的方法进行求解。最原始的方法是,穷举 α 、 β 和 m 之值,对于每一种组合,将时间数据代入Weibull函数从而计算出药物累计溶出百分比的估计值,用实际值与估计值计算出 e_i ,从中找出使得 e_i 平方和最小的一种组合。这种方法可以利用计算机语言编写程序进行求解。另外,还可以构造一个非

线性规划的数学模型,利用现有软件如MATLAB、Lingo等进行计算(本文不再赘述):

$$\min z = \sum_{i=1}^n e_i^2$$

$$\text{st. } e_i = F_i - (1 - e^{-\frac{(t_i - \alpha)^m}{\beta}}) \quad i = 1, 2, \dots, n$$

$$\alpha \geq 0.01$$

$$\alpha \leq t_i - 0.01$$

其中, F_i 、 t_i 、 n 均为已知数据, z 为目标函数值, st.指约束方程。

综上所述,在用OLS方法进行Weibull参数估计时,要进行自相关性、异方差性检验,如不能通过检验,则要另辟蹊径。实际上,不仅是在进行Weibull参数的估计,还是在其他有关模型参数的估计,只要是采用了OLS,就必须进行自相关性、异方差性检验。

参考文献

- [1] 史文欣,张琦岩.片剂溶出度测定数据处理的电子计算机程序[J].沈阳药学院学报,1987,4(1):5.
- [2] 张莉,夏运岳.电子表格Excel计算药物溶出度Weibull分布参数[J].药学进展,2002,26(1):48.
- [3] 吴爱琴,赖宝龙,任斌,等.格列吡嗪片体外溶出度比较[J].广东药学,2005,15(3):60.
- [4] 邱家学.灰色系统理论在求解片剂溶出度Weibull分布参数中的应用[J].中国药科大学学报,1988,19(1):17.
- [5] 何晓群,刘文卿.应用回归分析[M].3版.北京:中国人民大学出版社,2011:6.
- [6] 达莫达尔.经济计量学精要[M].张涛,译.1版.北京:机械工业出版社,2000:225-230,247-252.
- [7] 古扎拉蒂.计量经济学:下册[M].林少宫,译.1版.北京:中国人民大学出版社,2000:407-416.
- [8] 古扎拉蒂.计量经济学:上册[M].林少宫,译.1版.北京:中国人民大学出版社,2000:358-369.
- [9] 尚北城,徐贵丽,张青,等.酮康唑微囊的制备及其药剂学研究[J].西南国防医药,2004,14(3):251.
- [10] 黄献,刘裕恒,莫志江.用SPSS拟合药物溶出度Weibull参数[J].中国药房,2006,17(14):1 079.
- [11] 易丹辉.数据分析与EViews应用[M].1版.北京:中国人民大学出版社,2008:30-83.
- [12] 刘会灯,朱飞.MATLAB编程基础与典型应用[M].1版.北京:人民邮电出版社,2008:194-238.
- [13] 洪楠,侯军.MINITAB统计分析教程[M].1版.北京:电子工业出版社,2008:114-144.
- [14] 茆诗松.统计手册[M].1版.北京:科学出版社,2003:492-493.
- [15] 李子奈.计量经济学[M].1版.北京:高等教育出版社,2003:57.
- [16] 白雪梅.异方差性的检验方法及评述[J].东北财经大学学报,2002(6):6.
- [17] 高辉.异方差性的诊断及数据属性影响的实证分析[J].山西师范大学学报:自然科学版,2003,17(4):1.

(收稿日期:2012-02-09 修回日期:2012-04-05)